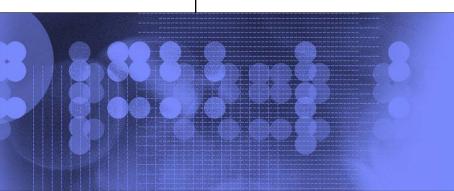




Data I/O Research at IBM

Roger Haskin, Dilip Kandlur, Julian Satran, T Basil Smith





Research and Development that IBM is engaged in now

Parallel file systems (GPFS)

- Multi-cluster
- Scalability
- Support for object storage
- Storage management (ILM)

NFSv4

- Clustered NFSv4
- Global namespace
- pNFS

Archival Storage

- HPSS

New Storage Technologies

- Object storage
- Reliable storage (e.g. advanced RAID)
- Ice Cube
- Solid-state memory for storage
- Content search



What research should be emphasized

Hierarchical storage

- Tape is too slow (latency and throughput) and expensive (robotics)
- Not practical to back up large file systems
- New hardware technology indicated, but SATA still too expensive

Global file systems

- Researchers need to share data to deliver on the promise of grid computing
- Solutions needed across a spectrum of backbone configurations
 - Conventional IP networks
 - Dedicated multi-gigabit WANs
- Integrity and privacy/anonymity
- Access control

Management

- Systems getting too large and complex
- Difficult to plan, deploy, do problem determination, measure performance
- Data management for large data sets
- Open standards needed

Content management and search



Areas that need more focus

- Searching and indexing
- Characterization (metadata, attributes) for non-structured information (IBM is developing a framework that it intends to make open)
- Preservation in terms more general than the archive document data
- Standards what higher level standards should we pursue?



Shift emphasis/Different focus

- To make low cost/high quality products widely available manufacturers need volume
- The market for very high performance data I/O is limited
- Standards (especially stable standards) encourage manufacturers
- Departure from conventional widely used technology is hard and often made harder by an attempt to fit the needs of every sector of the market. The introduction of some widely sought technology becomes unlikely (e.g., Object Storage)
- Academia and the open-source community might ease the pain of introducing new technology by adopting it into well understood widely used applications and subsystems (e.g., Object Storage)